

Israel de Sousa Brandão Edmundo

Identificação de padrões de ocorrência de incêndios no Parque Nacional da Serra da Canastra a partir de mineração de dados

XV Curso de Especialização em Geoprocessamento  
2014



UFMG  
Instituto de Geociências  
Departamento de Cartografia  
Av. Antônio Carlos, 6627 – Pampulha  
Belo Horizonte  
cartog@igc.ufmg.br

**ISRAEL DE SOUSA BRANDÃO EDMUNDO**

**IDENTIFICAÇÃO DE PADRÕES DE OCORRÊNCIA DE INCÊNDIOS NO PARQUE  
NACIONAL DA SERRA DA CANASTRA A PARTIR DE MINERAÇÃO DE DADOS**

Monografia apresentada como requisito à obtenção do grau de especialista em Geoprocessamento. Curso de Especialização em Geoprocessamento. Departamento de Cartografia. Instituto de Geociências. Universidade Federal de Minas Gerais.

Orientadora: Prof<sup>ª</sup> Dra. Maria Márcia Magela Machado

Co-orientadora: M<sup>ª</sup>. Amanda Alves dos Santos

**BELO HORIZONTE**

**2014**

Israel de Sousa Brandão Edmundo

Monografia defendida e aprovada em cumprimento ao requisito exigido para obtenção do título de Especialista em Geoprocessamento, em 03 de dezembro de 2014, pela Banca Examinadora constituída pelos professores:

---

Profª Dra. Maria Márcia Magela Machado

---

Mª. Amanda Alves dos Santos

---

Profª Dra. Úrsula Ruchkys de Azevedo

## RESUMO

Incêndios florestais ocorrem por motivos variados, seja pela ação do homem ou pelas condições climáticas, e provocam prejuízos ambientais e financeiros. Desta forma, aumentar a eficiência do combate a incêndios florestais é sempre um desafio e uma necessidade. Este trabalho tem como objetivo a identificação padrões de ocorrência de incêndios florestais utilizando técnicas geoprocessamento e de mineração de dados no Parque Nacional da Serra da Canastra. Foram coletadas informações de focos de calor no parque no período de 2000 a 2014 e utilizado variáveis fisiográficas de uso e ocupação do solo, declividade e orientação da encosta e de ignição como distância das estradas e trilhas e atrativos do parque. Essas informações foram sobrepostas aos focos de calor e extraído pontualmente os dados relativos a cada uma das variáveis. Esses foram submetidos ao processo de mineração de dados gerando padrões de ocorrência que foram replicados para as demais áreas. Depois de identificado os padrões, foi gerado um mapa potencial de ocorrência de incêndio da área com base nas informações da árvore de decisão gerada. O resultado do trabalho foi satisfatório e apresentou padrão de potencialidades de ocorrência de incêndios interessante, levando em consideração que as áreas classificadas como um alto potencial não seguiu um padrão característico de incêndios florestais.

Palavras chaves: Incêndios Florestais, Mineração de Dados, Análise Multicritérios

## SUMÁRIO

Pág.

<b>1. INTRODUÇÃO .....</b>	<b>8</b>
<b>2. FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>10</b>
2.1 – Incêndios florestais .....	10
2.2 – Prevenção de incêndios florestais e geoprocessamento .....	11
2.3 – Mineração de dados.....	12
<b>3. CARACTERIZAÇÃO DA ÁREA DE ESTUDOS .....</b>	<b>15</b>
<b>4. MATERIAIS E MÉTODOS .....</b>	<b>17</b>
4.1 – Materiais.....	17
4.2 – Métodos .....	17
4.2.1 Seleção das variáveis .....	17
4.2.2 Aquisição dos dados.....	17
4.2.3 Tratamento dos dados.....	18
4.2.4 Extração das informações .....	23
4.2.5 Minerador de dados .....	23
4.2.6 Elaboração do mapa final .....	24
<b>RESULTADOS E DISCUSSÕES .....</b>	<b>25</b>
5.1. Interpretação das estatísticas .....	25
5.2. Mapa potencial.....	26
<b>5. CONCLUSÕES.....</b>	<b>28</b>
<b>6. REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>29</b>
<b>ANEXO 1 – ARVORE DE DECISÃO.....</b>	<b>32</b>
<b>ANEXO 2 – MAPA DE POTENCIAL .....</b>	<b>36</b>

## LISTA DE FIGURAS

	<u>Pág.</u>
Figura 1 – Etapas do processo de KDD adaptado de FAYYAD.....	13
Figura 2 - Localização da área de estudos .....	15
Figura 3 - Espacialização dos focos de calor.....	19
Figura 4 – Densidade dos focos de calor .....	20
Figura 5 - Mapa de declividade da área.....	21
Figura 6 - Mapa de orientação da vertente .....	21
Figura 7 - Mapa de trilhas e atrativos do parque .....	22
Figura 8 - Mapa de uso e ocupação .....	23
Figura 9 - Arquivo de entrada padronizado para o WEKA.....	24
Figura 10 - Índice Kappa .....	25
Figura 11 - Razão entre os dados do modelo e o erro .....	26
Figura 12 - Erro quadrático médio .....	26
Figura 13 - Mapa potencial de incêndios.....	27

## LISTA DE TABELAS

Pág.

Tabela 1 - Classificação do índice kappa ,daptado de Landis e Koch, 1977 .....	25
--	----

## LISTA DE SIGLAS E ABREVIATURAS

CIMAN	- Centro Integrado Multiagencias de Coordenação Operacional
CODEMIG	- Companhia de Desenvolvimento Econômico de Minas Gerais
CPRM	- Companhia de Pesquisa de Recursos Minerais
DEM	- Digital Elevation Model
IBAMA	- Instituto Brasileiro de Meio Ambiente
ICMBio	- Instituto Chico Mendes de Conservação da Biodiversidade
INPE	- Instituto Nacional de Pesquisas Espaciais
KDD	- Knowledge Discovery in Database
NASA	- National Aeronautics and Space Administration (Administração de Aeronáutica e Espaço dos EUA)
PNSC	- Parque Nacional da Serra da Canastra
PREVFOGO	- Sistema Nacional de Prevenção e Combate aos Incêndios Florestais
SAD	- South American Datum
SIG	- Sistema de Informação Geográfica
UC	- Unidade de Conservação
UTM	- Universal Transverso de Mercator (Projeção Cartográfica)
WGS84	- World Geodetic System 1984 (Sistema Geodésico Mundial de 1984)



## 1. INTRODUÇÃO

Atualmente, um dos maiores desafios na gestão de Unidades de Conservação (UC) do Brasil é a prevenção e o combate aos incêndios florestais. Esses são, na sua grande maioria, causados por ações antrópicas direta ou indiretamente. Entretanto, elementos climáticos como a estiagem, baixa umidade, a velocidade do vento e também o relevo do local são fatores que influenciam na sua propagação.

O fogo em áreas de UC ou em suas proximidades pode ser causado intencionalmente pelo homem com diferentes fins como: a renovação de área de pastagem, limpeza de área para cultivo, extrativismo vegetal, a queima de lixo entre outros. Porém, muitas das vezes apesar das restrições legais, esse tipo de queimada é inconsequente, já que geralmente não se planeja um controle para que o fogo seja extinto assim que se atingir o objetivo.

Estudos sobre o impacto do fogo em florestas tropicais tornam-se cada vez mais importantes à medida que os efeitos provocados pelos incêndios passam a ter repercussões globais negativas, particularmente sobre a atmosfera e o estoque de biodiversidade (COCHRANE, 2003).

Na Serra da Canastra, os registros de ocorrências de incêndios são bastante antigos, sendo descritos por naturalistas ainda século XVI. Por ser uma área de campo muito aberto, ela é frequentemente atingida por raios, e como a região apresenta alto número de propriedades rurais, os incêndios criminosos provocados pelos proprietários dessas terras são bastante frequentes.

Uma das melhores formas para amenizar os danos causados pelo fogo é ter um conhecimento prévio de quais são as áreas de maior probabilidade para que se dê início um incêndio florestal. Felizmente há hoje no país uma crescente preocupação em relação aos recorrentes problemas causados pelo fogo em áreas de Unidades de Conservação e florestas em geral e com isso vem incentivando o desenvolvimento e aperfeiçoamento de técnicas preventivas e de combate, o que é imprescindível para a conservação de nossas paisagens naturais. As ações de prevenção e o combate aos incêndios florestais demandam um planejamento detalhado que contemple estratégias para a vigilância da área, medidas preventivas, recursos para aquisição e manutenção de equipamentos e ferramentas, remuneração de mão de obra, e formação de brigadas (PREVFOGO/IBAMA, 2005).

Os Sistemas de Informação Geográfica (SIG), são uma boa alternativa para auxiliar no entendimento espacial do ambiente, pois eles retratam ou representam de forma mais fidedigna os problemas relacionados ao espaço, e desta forma são pontuais nas análises, podendo, no mais, construir uma diversidade de materiais cartográficos. (CERQUEIRA, 2006).

Assim, a elaboração de metodologias específicas de previsão se faz necessário para o suporte no planejamento do combate ao fogo, como a utilização de técnicas de geoprocessamento e a mineração de dados que podem ser utilizadas para identificação de padrões de ocorrência de incêndios, auxiliando na gestão de áreas naturais protegidas.

Este trabalho teve como objetivo geral identificar áreas potenciais de ocorrência de incêndios no Parque Nacional da Serra da Canastra (PNSC), através da análise de fatores naturais e antrópicos de forma a proporcionar o suporte ao planejamento e ação no combate a incêndios nesta unidade de conservação. Neste contexto, o objetivo específico é gerar um mapa do PNSC especificando o potencial, em três classes: alto, médio e baixo, de ocorrência de incêndios florestais.

## **2. FUNDAMENTAÇÃO TEÓRICA**

### **2.1 – Incêndios florestais**

O fogo é um dos principais responsáveis por problemas ambientais, econômicos e mesmo humanos no Brasil e no mundo (SOARES e SANTOS, 2002). O incêndio florestal é definido pelo IBAMA (2011) como sendo os efeitos causados pelo fogo em sua livre propagação por meio dos processos de transmissão de calor, sem limites pré-estabelecidos, sob a incitação das forças atuantes de clima, relevo e combustível em biomassa vegetal que podem provocar prejuízos importantes ao ambiente, como os danos a fauna, a flora e às pessoas, inclusive com perdas de vidas. Ele também pode acarretar algumas consequências econômicas consideráveis, como a destruição de habitat e os custos para controlar o fogo (SOARES e SANTOS, 2002).

A falta de chuva, a baixa umidade relativa do ar e a vegetação seca, proporcionam um quadro ideal para que ocorra com frequência incêndios florestais. O vento auxilia na combustão uma vez que o fogo é iniciado, fornecendo oxigênio extra e também na direção de propagação do mesmo como afirma Brown e Davis (1973).

Bontempo (2006) afirma que a importância do relevo se dá por meio da altitude, exposição e inclinação do terreno que também tem uma forte influência, uma vez que se relaciona com o clima e a vegetação. No Estado de Minas Gerais, muitos parques estão localizados em altitudes elevadas e relevos bastante acidentados, o que faz essas regiões, além de outros fatores, propícias a ocorrência de incêndios florestais.

Associados aos aspectos físicos e bióticos as diversas atividades antrópicas presentes nas proximidades da vegetação influenciam, direta ou indiretamente no aumento dos incêndios florestais ao longo dos anos, porém, segundo Soares e Santos (2002), são os fatores climáticos, como seca e velocidade do vento, ou o relevo do local que influenciam na sua propagação e determinam os seus efeitos devastadores.

Esses mesmos autores apontam que, saber a causa dos incêndios permite trabalhar objetivamente com a prevenção e redução das fontes de fogo, otimizando os gastos com a proteção das florestas. Ainda é possível destacar que os danos econômicos podem ser estimados a partir da identificação do tipo de vegetação atingida e que a eficiência do combate pode ser medida por meio da classificação dos incêndios por classe de tamanho.

Batista (1990) descreve que na floresta pode ocorrer quatro tipos de incêndios: Subterrâneo ou de solo, de superfície, de copa e de projeção.

Nos incêndios subterrâneos ou de solo, o fogo queima abaixo da superfície onde está localizada a matéria orgânica acumulada. Os incêndios de superfície são o tipo mais comum ocorrendo a queima dos combustíveis depositados no solo da floresta inicialmente como gramíneas, ramos e folhas. Esse tipo de incêndio pode dar origem aos incêndios de copa se não forem combatidos com rapidez ou se as condições climáticas estiverem favoráveis. Já os incêndios de copa atingem os materiais combustíveis acima de 1,80m de altura, geralmente é iniciado por um incêndio superficial, com exceção dos provocados por raios ocorrendo principalmente em dias de muito vento e com baixa umidade relativa do ar. O seu combate é extremamente complicado e destrutivo, pois ocorre nas raízes da vegetação, de forma lenta e sem chamas. No Brasil, esse tipo de incêndio é pouco comum.

O ultimo tipo de incêndio, projeção, é devido ao carregamento de material incandescente para outras regiões na maioria das vezes provocada pelo vento.

Possuir o conhecimento prévio das áreas de maior probabilidade que ocorra um incêndio é uma etapa importante para que os danos sejam minimizados, assim, Nogueira (2002) ressalta que a vigilância e a fiscalização são fatores essenciais no planejamento da proteção florestal para a redução dos casos de incêndio uma vez que a melhor forma de se combater um incêndio florestal é através de ações preventivas.

## **2.2 – Prevenção de incêndios florestais e geoprocessamento**

Para dar suporte ao combate de incêndios florestais e queimadas no Brasil, o governo federal criou em 1989, por meio do Decreto nº 97.635, o Sistema Nacional de Prevenção e Combate aos Incêndios Florestais – PREVFOGO, que atribuiu ao ICMBio (Instituto Chico Mendes de Conservação da Biodiversidade) a competência de coordenar as ações necessárias à organização, implementação e operacionalização das atividades relacionadas à educação, pesquisa, prevenção e combate aos incêndios florestais e queimadas.

No dia 15 de outubro de 2014, foi lançado pelo ICMBio, em parceria com o Instituto Nacional de Pesquisas Espaciais (INPE), o Centro Integrado Multiagências de Coordenação Operacional (Ciman). Trata-se de uma plataforma virtual criada para monitorar as queimadas

e focos de incêndio em todo o país, o Ciman tem como objetivo disponibilizar informações em tempo real dos locais com focos de calor e a localização das equipes de brigadistas ou bombeiros que estiverem em campo no combate ao fogo.

Com o desenvolvimento das geotecnologias ao longo dos últimos, o planejamento e prevenção de incêndios florestais esta cada vez mais eficiente onde o suporte proporciona um atendimento mais rápido além de gerar uma economia operacional como um todo. Alguns satélites como o NOAA, EOS, GOES e METEOSAT possui sensores a bordo capazes de detectar focos de calor. Essas informações são fornecidas pelo site do INPE e utilizadas para monitorar queimadas.

Em trabalho realizado por Silva (2006), foi desenvolvida uma metodologia para analisar a susceptibilidade a incêndios florestais por meio da utilização do processamento digital de imagem, análise espacial e estatística *Analytical Hierarchy Process* (AHP). Utilizando dados de umidade, altimetria, dados meteorológicos e focos de calor foi possível caracterizar as ocorrências de incêndio naquela região por períodos do ano e características do relevo. Já no trabalho realizado por Prudente (2010), foi utilizado uma metodologia que aborda a atribuição de notas as variáveis utilizadas no modelo para que em seguida fosse realizada a álgebra de mapas para a criação de situações de risco de incêndio facilitando a tomada de decisão.

Nesse contexto, os Sistemas de Informações Geográficas (SIG) propiciam a criação de um banco de dados espacialmente consistente onde é possível realizar cruzamentos de um grande número de informações além de um acompanhamento temporal. Como ferramenta de análise e suporte dessas grandes quantidades de informações existe ferramentas e técnicas específicas, como o *data mining* ou mineração de dados.

### **2.3 – Mineração de dados**

Quando se tem uma grande quantidade de dados e é preciso que seja identificado padrões neste conjunto de informações, uma alternativa é utilizar um processo conhecido como Knowledge Discovery in Databases (KDD) e/ou minerador de dados.

As definições dos termos KDD e Data Mining ainda não possuem um consenso entre alguns autores que abordam o assunto. Rezende (2005), Wang (2005) e Han et al (2006) dizem que os termos são sinônimos, mas em contrapartida, para Cios et al (2007) e Fayyad (1966) o KDD

refere-se a todo o processo de extração do conhecimento de um dado inicial sendo o Data Mining apenas uma parte do processo como mostra a Figura 1.

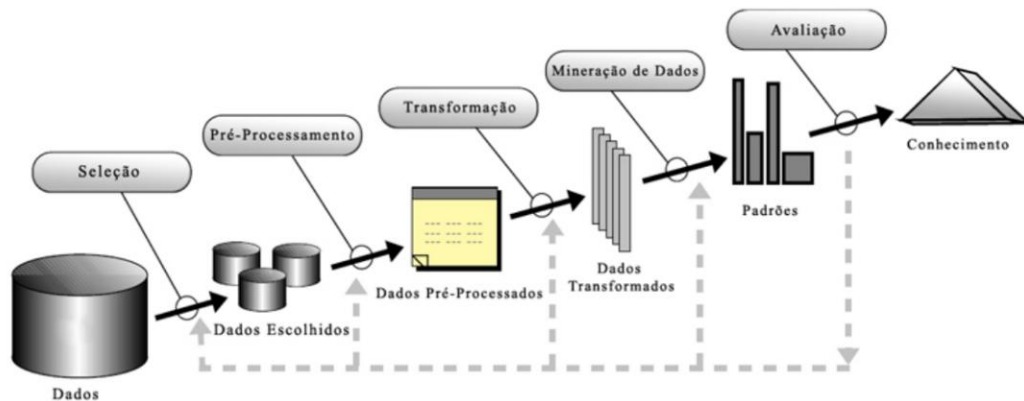


Figura 1: Etapas do processo de KDD adaptado de FAYYAD( 1996)

A mineração de dados envolve o uso de algoritmos para a compreensão inteligente dos dados e extração dos padrões úteis na produção de um modelo de conhecimento (FAYYAD et al., 1996), o que possibilita saber se fatores relevantes para o entendimento do fenômeno a ser trabalhado. Segundo Addrians e Zantinge (1996), a mineração de dados surge com base em três áreas: estatística clássica, inteligência artificial e aprendizado de máquina, sendo a primeira a mais antiga delas.

Ter uma base de dados confiável é um fator chave para o desenvolvimento do processo de mineração de dados e conseqüente a obtenção de resultados consistentes. O pré-processamento deve ser feito de forma criteriosa e com bastante atenção pois, segundo Mannila (1996), esta etapa, devido às dificuldades de integração de bases de dados heterogêneos, pode tomar até 80% do tempo de todo o processo.

Após ser feito o pré-processamento dos dados entra-se na etapa de transformação onde as informações devem ser padronizadas de acordo com o *software* que será utilizado para minerar os dados. Os *softwares* disponíveis para mineração de dados possuem uma grande quantidade de algoritmos de aprendizado, sendo necessário definir o algoritmo que irá proporcionar uma melhor resposta para o tipo de problema proposto. Sferra e Corrêa (2003) destacam que cada técnica de *Data Mining* utilizada para conduzir o processo adapta-se melhor a alguns problemas do que a outros, portanto, a experiência e intuição do analista

estão diretamente relacionadas ao bom resultado final depois do processo de mineração dos dados.

O pós-processamento é realizado através das etapas de avaliação e apresentação dos padrões, que são responsáveis pela identificação e análise dos padrões interessantes que representam conhecimento, bem como, a forma como o conhecimento extraído será apresentado ao usuário.

### 3. CARACTERIZAÇÃO DA ÁREA DE ESTUDOS

O Parque Nacional da Serra da Canastra (PNSC) está situado na região sudoeste do Estado de Minas Gerais e sua área decretada está situada em partes dos municípios de São Roque de Minas, Sacramento, Delfinópolis, Vargem Bonita, Capitólio e São João Batista do Glória. O parque foi criado pelo Decreto nº 70.355, datado de 3 de abril de 1972, com 200 mil hectares e preserva a nascente do rio São Francisco. Desta área total decretada, apenas 71.525 hectares estão sob o domínio de responsabilidade do ICMBio, enquanto que o restante ainda não é regularizado e são constituídos por propriedades privadas como aprestando na Figura 2. Para este trabalho, foi utilizada a área decretada total do parque.

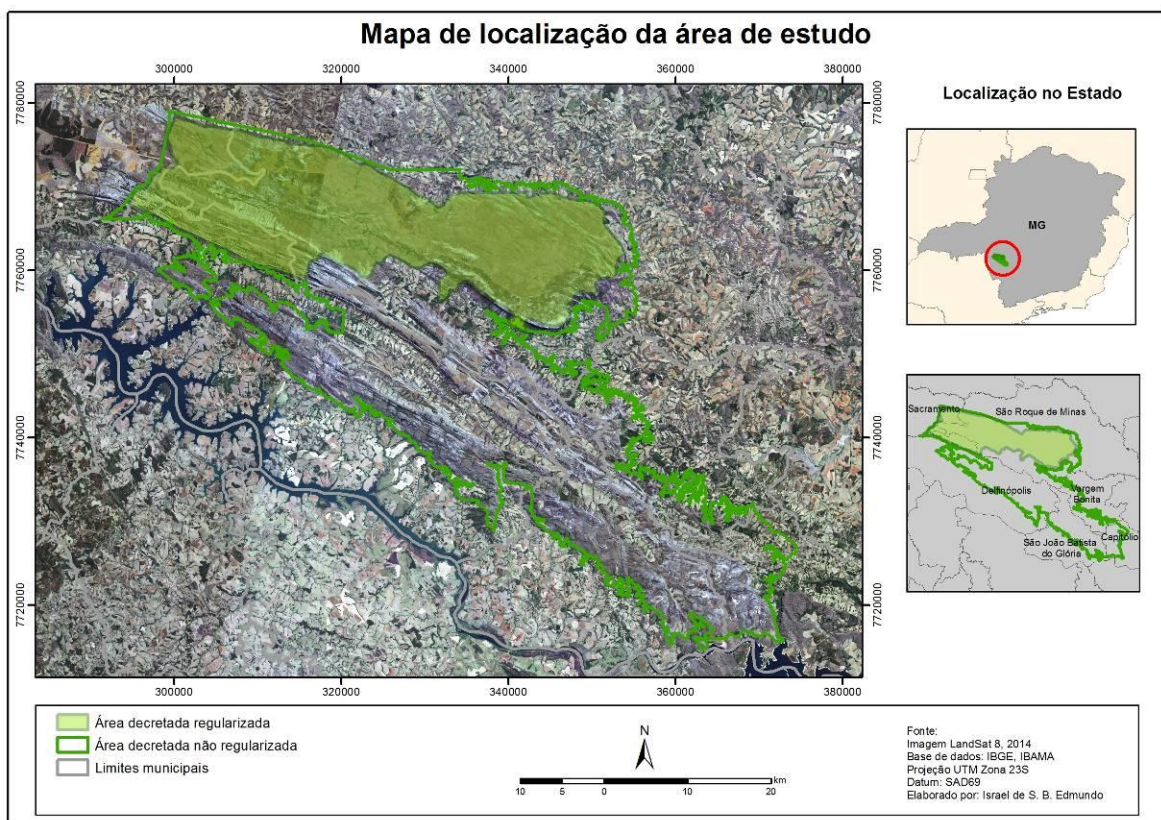


Figura 2: Localização da área de estudo

A região do PNSC está inserida no domínio fitofisionômico cerrado e sua vegetação consiste em florestas mesófilas de encosta, capões, cerradão, cerrado sensu stricto, campo cerrado, campo limpo e campo rupestre.

O clima predominante na região é o tropical sazonal com duas estações bem distintas: chuvosa durante o período de verão e inverno seco. A temperatura média anual fica em torno de 22-23° C e as máximas absolutas mensais não variam muito ao longo dos meses do ano,



podendo chegar a mais de 40° C. Em geral, a precipitação média anual fica entre 1.200 e 1.800 mm.

O relevo do PNSC se caracteriza por uma alternância de platôs, vales encaixados e encostas escarpadas podendo identificar a presença de dois grandes segmentos. O primeiro segmento, o maciço da Canastra é formado por um grande platô e possui uma vegetação predominantemente campestre. O segundo segmento possui um relevo mais acidentado e corresponde à região da chapada da Babilônia caracterizada pela alternância de faixas mais estreitas desta feição a encostas escarpadas e vales alongados, sempre com orientação NW-SE, seguindo o controle dos maciços de quartzito.

Segundo o plano de manejo (IBAMA, 2005) dessa unidade de conservação, o parque possuiu um significativo histórico de incêndios. Esses são registrados, para o chapadão da Canastra, de forma continuada desde 1987 possibilitando sua análise e caracterização. De acordo com esse documento, a principal causa de incêndios no parque no período de 1987 a 2001 foi a ação antrópica, aproximadamente 50% dos 87 incêndios registrados. A porcentagem restante é causada por raios (44%) e por causas desconhecidas (6%). Do total de incêndios causados pela ação antrópica, nesse período, o plano de manejo ainda aponta que 47% incendiários, 23% com queima de pastagem e 30% não tem causas definidas.

## 4. MATERIAIS E MÉTODOS

### 4.1 – Materiais

Para esta pesquisa foram utilizadas as seguintes bases cartográficas e softwares:

Dados vetoriais:

- Uso e ocupação do solo do PNSC – SANTOS(2014) - 1:150.000
- Atrativos atuais do PNSC - CODEMIG/CPRM - 1:1.000.000
- Estradas e trilhas do PNSC – Vetorizado através de imagens de satélite
- Área decretada e regulamentada – ICMBio/MMA
- Limite do PNSC (Área decretada) – ICMBio/MMA

Dado matricial:

- Imagem GDEM ASTER, United States Geological Survey

Dados alfanuméricos:

- Coordenadas dos focos de calor – INPE

Softwares:

- ArcGis 10.2, licença tipo ArcInfo
- WEKA Explore 3.6.9

### 4.2 – Métodos

#### 4.2.1 Seleção das variáveis

As variáveis selecionadas para serem utilizadas neste trabalho foram divididas em duas categorias: fatores fisiográficos e fatores de ignição.

Para os fatores fisiográficos foram utilizados a declividade, o uso e ocupação do solo e a orientação da encosta, enquanto que para os fatores de ignição foram utilizados as estradas que cortam o parque, as trilhas e atrativos presentes no parque, e os focos de incêndio ocorridos no período pré-determinado.

#### 4.2.2 Aquisição dos dados

As coordenadas de focos de calor do parque foram adquiridas no site da Divisão de Processamento de Imagens do Instituto Nacional de Pesquisas Espaciais. Estes dados fazem

parte do Sistema de Monitoramento de Queimadas e Incêndios e são atualizados de 3 em 3 horas diariamente. Os dados trabalhados representam os focos de calor captados pelo satélite AQUA com sensor Modis que é utilizado como satélite referência<sup>1</sup> desde 2011 e do satélite NOAA que era utilizado como satélite de referência antes de 2011. O período de coleta considerado foi de 01 Janeiro de 2000 a 31 de dezembro de 2014. Foram registrados 949 focos de calor dentro do limite decretado do parque. Esses dados foram reprojctados para a projeção plana UTM datum SAD69 para a sua padronização.

As imagens GDEM ASTER são disponibilizadas gratuitamente pelo United States Geological Survey (USGS) na resolução espacial de 30m.

As estradas e trilhas foram vetorizadas a partir de imagens de alta resolução obtidas através do *software* Google Earth Pro fixada a escala de 1:50.000

A base de uso e ocupação foi adquirida em formato vetorial da dissertação de Santos (2014), que utilizou imagens do satélite Landsat do ano de 2013 para elabora-la.

Os atrativos atuais do PNSC foram disponibilizados pela CODEMIG/CPRM na escala de 1:1.000.000 em formato vetorial de ponto.

O limite do parque e a área regularizada são disponibilizadas de forma gratuita e em formato vetorial no site do ICMBio/MMA.

#### ***4.2.3 Tratamento dos dados***

Os focos de calor disponibilizados para a área do parque se encontram em formato alfanumérico com coordenadas geográficas em formato de grau decimal. Essas coordenadas foram tratadas para dar entrada no *software* ArcGIS e serem espacializadas em formato vetorial georreferenciado como apresentado na Figura 3.

---

<sup>1</sup> O satélite de referência é aquele cujos dados de focos de queima de vegetação são utilizados para comparações temporais no monitoramento do INPE

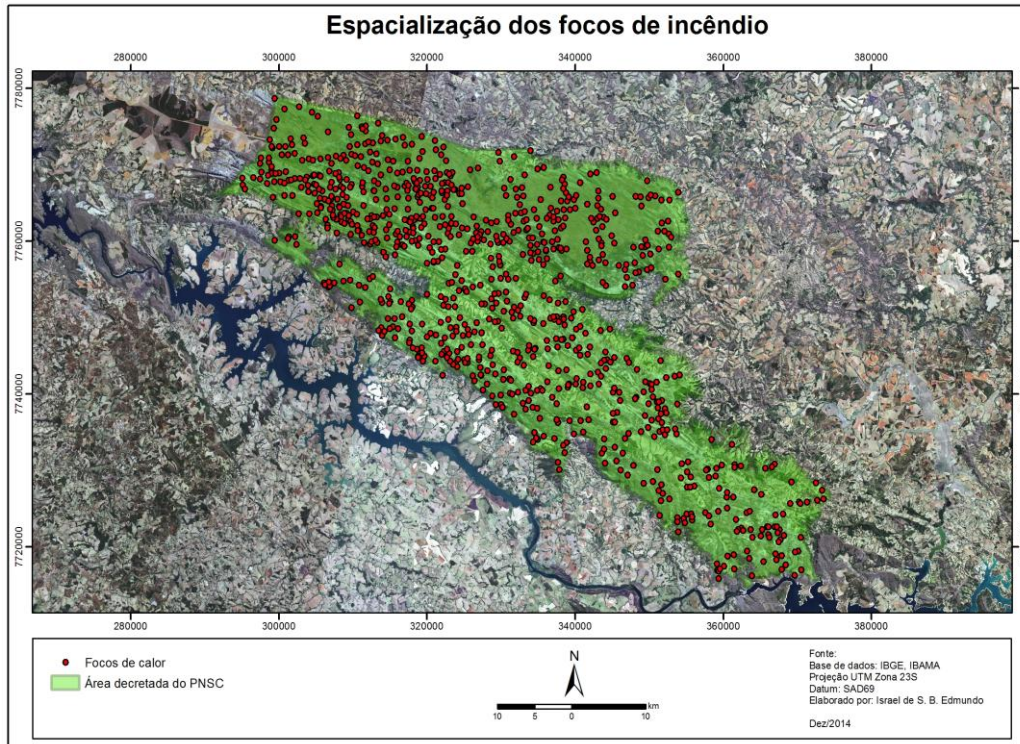


Figura 3: Espacialização dos focos de calor

A densidade dos focos de calor foi obtida através da ferramenta kernel density também do software ArcGis, onde foi considerado o valor de 2000m como raio de pesquisa no campo específico solicitado pelo software. Essa matriz então foi classificada em 3 classes, alta, média e baixa densidade (Figura 4).



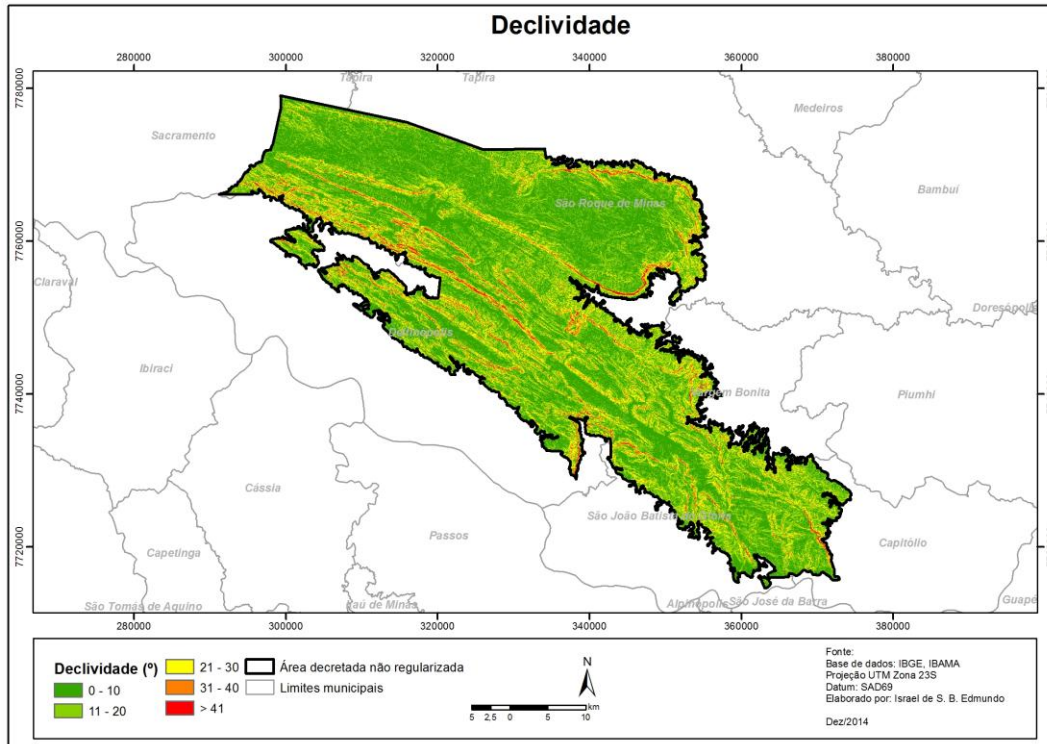


Figura 5: Mapa de declividade da área

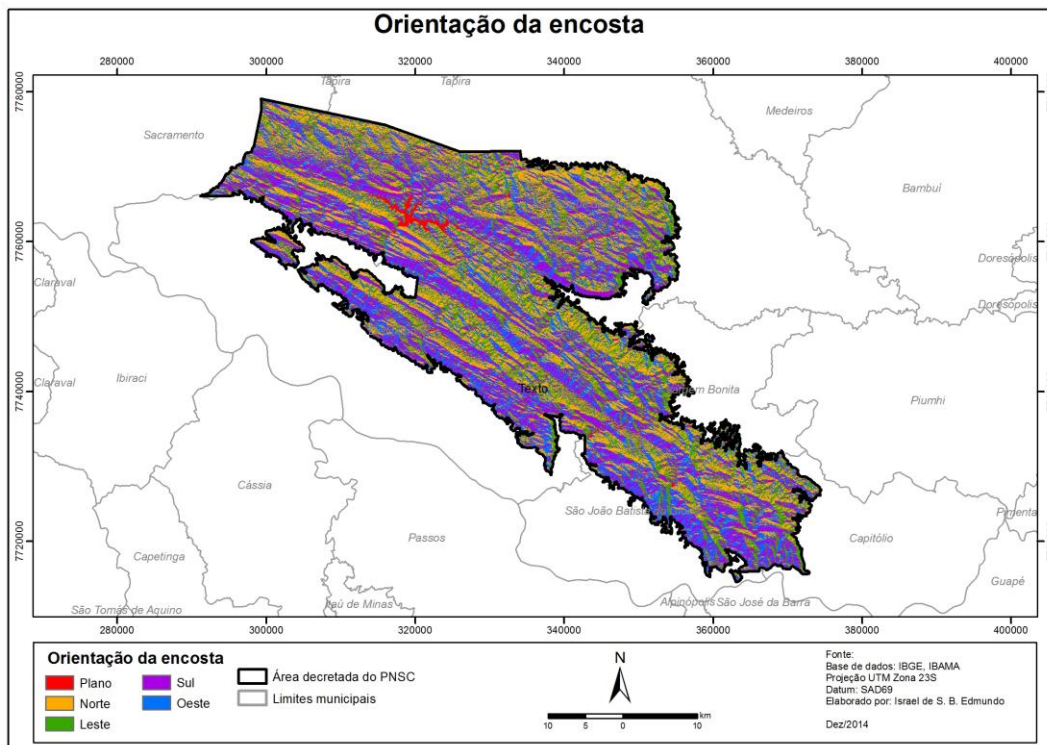


Figura 6: Mapa de orientação da vertente

As camadas de trilhas, estradas e atrativos presentes no parque (Figura 7) foram utilizadas partindo do princípio de distância de influência de cada uma. Para isso foi gerada uma matriz de distância dessas camadas utilizando a ferramenta *Euclidian Distance* da extensão *Spatial*

*Analyst* que atribui para cada pixel o valor relativo à distância das camadas citadas em relação aos pontos de focos de incêndio.

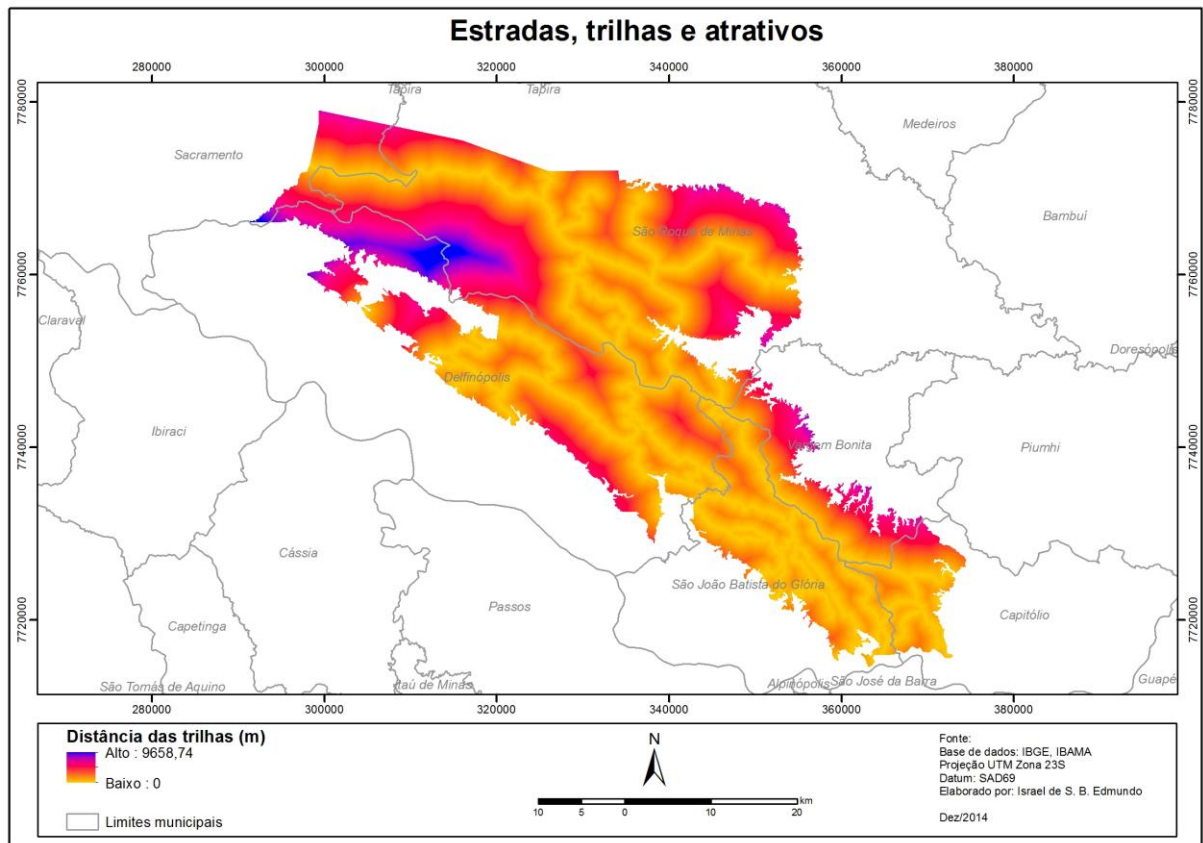


Figura 7: Mapa de trilhas e atrativos do parque

O mapa de uso e ocupação do solo (Figura 8), foi adquirido em formato vetorial e foi convertido para matriz através da ferramenta *Polygon to Raster* do *software* ArcGIS para que pudesse ser extraído os valores do pixel.

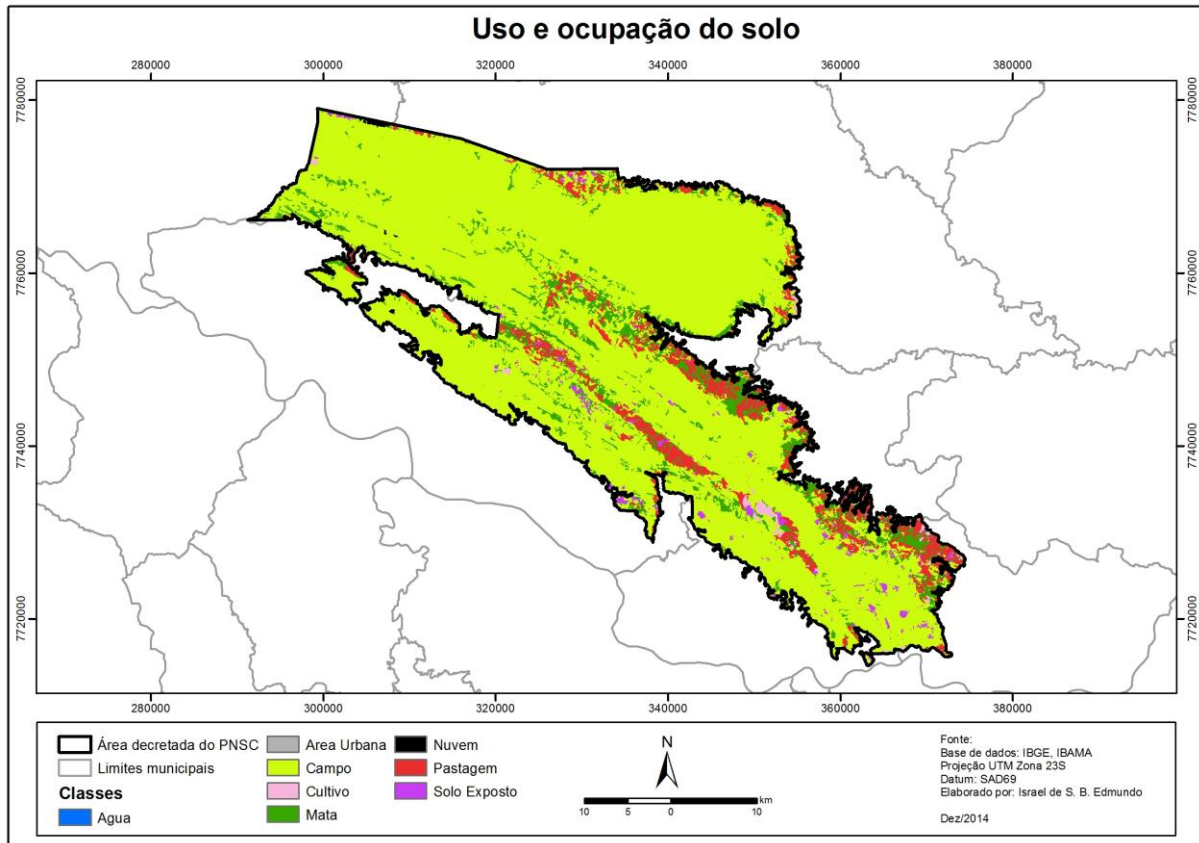


Figura 8: Mapa de uso e ocupação do solo

#### 4.2.4 Extração das informações

Depois de gerado as matrizes de informação de declividade, orientação da encosta, uso e ocupação, distância das trilhas, estradas e atrativos e densidade, foi necessário extrair as informações relacionadas ao pixel onde cada ponto marcado como focos de calor. Para este processo, foi utilizada a ferramenta *Extract multi values to point* onde o dado de entrada são os pontos de focos de calor e os rasters utilizados para a análise. O resultado do processo implica na extração dos atributos dos rasters de entrada pontualmente.

#### 4.2.5 Minerador de dados

Visando encontrar padrões nas ocorrências de incêndio foi utilizado o *software WEKA* para realizar a mineração dos dados e o classificador por árvore de decisão com o algoritmo J48. O algoritmo J48 é considerado o mais popular algoritmo do WEKA. Ele é uma implementação do algoritmo C4.5 release 8 que, gera árvore de decisão baseado em um conjunto de dados de treinamento.



Os dados extraídos dos *rasters* gerados anteriormente foram padronizados e organizados em um arquivo de texto no formato específico aceito pelo *software* como arquivo de entrada. A Figura 8 apresenta as primeiras linhas da combinação presente nesse arquivo.

Para as variáveis declividade, distância dos atrativos e distância das trilhas não foi atribuído ou pré-definido classes de entrada, pois desta maneira, utilizando o valor numérico real de cada pixel para cada uma dessas variáveis, o próprio minerador de dados interpreta e separa em classes que ele considera como sendo de maior importância.

```
@relation "incendio"

@ATTRIBUTE declividade NUMERIC
@ATTRIBUTE dist_atriv NUMERIC
@attribute dist_trilha NUMERIC
@attribute orient_encost {Plano, Norte, Sul, Leste, Oeste}
@attribute USO_OCUP {Campo, Mata, Solo_exp, Pastagem}
@attribute Densidade {Alta Media, Baixa}

@data
0.00,11394.40,7460.90,Plano,Campo,Alta
0.00,1601.00,1008.02,Plano,Campo,Media
0.00,9041.70,5436.63,Plano,Campo,Media
0.00,11031.60,7231.49,Plano,Campo,Alta
0.00,10326.70,6718.46,Plano,Campo,Alta
0.00,1351.33,2401.69,Plano,Campo,Baixa
0.00,5215.00,2550.00,Plano,Mata,Baixa
0.00,10543.50,6761.92,Plano,Campo,Alta
0.00,9931.27,6395.14,Plano,Campo,Alta
```

Figura 9: Arquivo de entrada padronizado para o WEKA

#### 4.2.6 Elaboração do mapa final

O mapa final de potencial foi elaborado por meio da espacialização das regras de associação de variáveis geradas pela árvore de decisão. Para esse processo, foi feita uma reclassificação das informações de cada nível da árvore gerada utilizando a ferramenta *Reclass* do *software* ArcGis, atribuindo 1 ao valor real que corresponde a variável e 0 para os demais de cada variável. Em seguida foi feita uma álgebra de mapa com os valores reclassificados para que se pudesse agrupar todos dos potenciais altos, médios e baixos de cada folha da árvore.

## RESULTADOS E DISCUSSÕES

### 5.1. Interpretação das estatísticas

A classificação feita pelo método de árvore de decisão permitiu a elaboração de um mapa potencial, através de álgebra de mapas, composto por três classes: Alto, Médio e Baixo.

O topo da árvore de decisão, ou nó raiz, é a variável considerada pelo classificador como a mais significativa dentre todas as outras utilizadas no modelo. Neste caso, o classificador considerou a variável distância das trilhas como com a maior relevância. A árvore de decisão resultante da mineração de dados pode é apresentada no Anexo 1.

O índice Kappa calculado a partir da mineração dos dados foi de 0,5719, que segundo, a tabela de classificação proposta por Landis e Koch (1977), é um resultado considerado como foi satisfatório. O índice Kappa varia de 0 ( uma classificação quase inexistente) e 1 ( uma classificação quase perfeita), como mostra a escala de Landis e Koch(1977), na Tabela 1.

Índice de Kappa	Classificação
<0,00	Quase inexistente
0,00 - 0,20	Pequena
0,21 - 0,40	Insatisfatória
0,41 - 0,60	Satisfatória
0,61 - 0,80	Substancial
0,81 - 1,00	Quase perfeita

Tabela 1: Classificação do indice kappa ,daptado de Landis e Koch, 1977

=== Summary ===

Correctly Classified Instances	693	73.0242 %
Incorrectly Classified Instances	256	26.9758 %
Kappa statistic	0.5719	
Mean absolute error	0.2437	
Root mean squared error	0.3491	
Relative absolute error	56.6065 %	
Root relative squared error	75.2411 %	
Total Number of Instances	949	

Figura 10: Índice kappa

A razão entre os dados do modelo e o erro (F-Measure) mede a precisão de cada classe da mineração. Quanto mais próximo de 1 mais preciso está a classificação do minerador e quanto mais próximo de 0 menor precisão. Conforme mostrado na Figura 11, as classes Alta e Média

obtiveram bons resultados, enquanto que para a classe Baixa o valor é considerado intermediário para bom. Nessa classe o valor resultante um pouco mais baixo indica que o *software* teve mais dificuldades para distinguir quais valores pertenciam a esta classe. O uso de mais variáveis no modelo poderia auxiliar na distinção das classes aumentando a precisão da mineração na classe baixa.

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.726	0.058	0.796	0.726	0.759	0.926	Alta
	0.864	0.313	0.686	0.864	0.765	0.861	Media
	0.549	0.074	0.777	0.549	0.644	0.868	Baixa
Weighted Avg.	0.73	0.176	0.741	0.73	0.725	0.879	

Figura 11: Razão entre os dados do modelo e o erro

O erro médio quadrático mostra se o modelo está ajustado aos dados inseridos na mineração, quanto mais próximo de 0, mais ajustado aos dados está o modelo. Neste trabalho, o erro médio quadrático calculado foi de 0,3491 conforme mostra a Figura 12. Sendo assim, podemos considerar que o modelo está ajustado.

=== Summary ===

Correctly Classified Instances	693	73.0242 %
Incorrectly Classified Instances	256	26.9758 %
Kappa statistic	0.5719	
Mean absolute error	0.2437	
Root mean squared error	0.3491	
Relative absolute error	56.6065 %	
Root relative squared error	75.2411 %	
Total Number of Instances	949	

Figura 12: Erro quadratico médio

## 5.2. Mapa potencial

O mapa de potencial de ocorrência de incêndios foi elaborado a partir dos dados resultantes da árvore de decisão como apresentado na Figura 13.

O mapa apresenta três classes, de acordo com o risco potencial para a ocorrência de incêndios. A classe Alta representa aproximadamente 12% da área, a classe Média corresponde a 61%, enquanto que a classe Baixa representa aproximadamente 27% da área total.

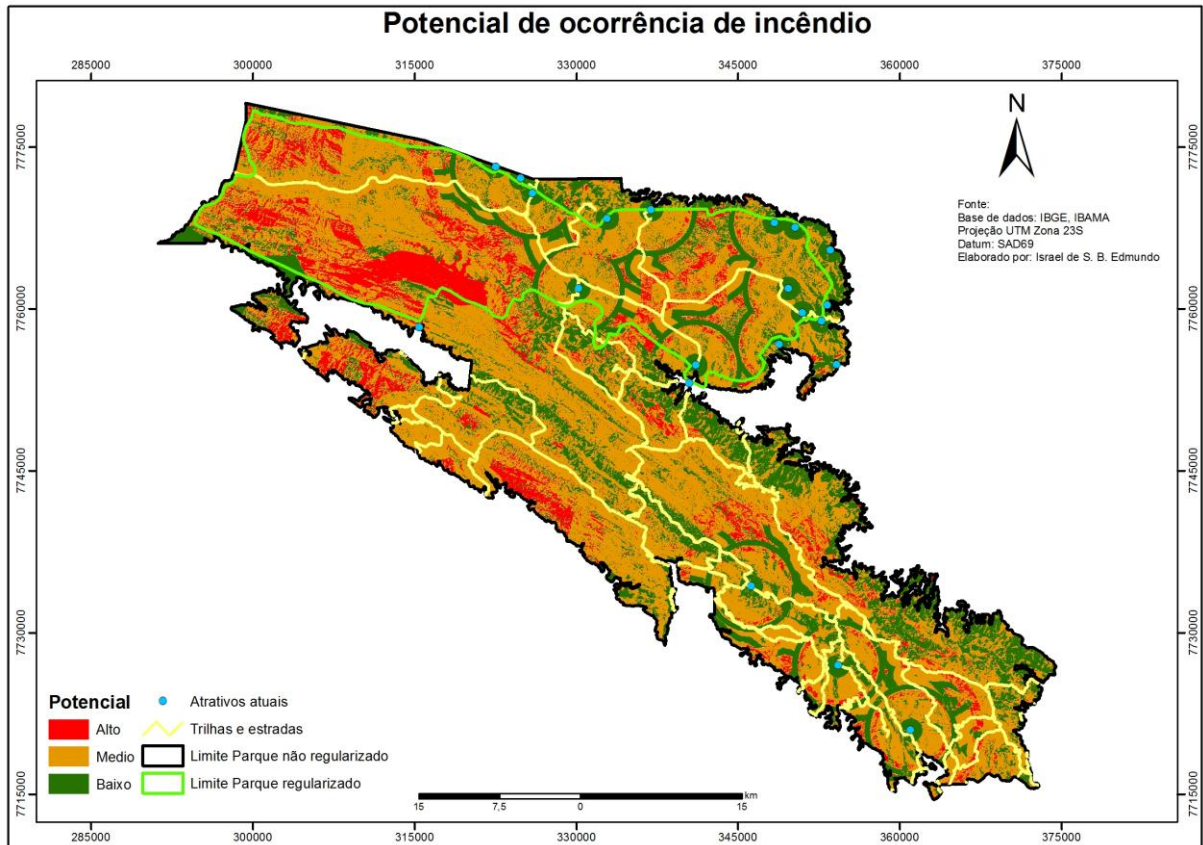


Figura 13: Mapa potencial de incêndios

Por meio da análise do mapa potencial, é possível observar que a região com o maior potencial de ocorrência de incêndio está localizada dentro do limite regularizado do parque, sendo que do total de 12% de áreas classificadas como alto potencial, aproximadamente 7% está localizado dentro desta área. Além disso, é possível observar que as manchas de alto potencial, não estão tão próximas das trilhas, estradas e estruturas do parque como seria esperado, o que poderia justificar o alto número registrado de ocorrência de incêndios iniciado por ação de raios conforme foi citado no capítulo 3 deste trabalho, por ser uma área de grande altitude e com a declividade pouco acentuada.

Apesar da porção sudeste do parque não regularizado possuir uma quantidade significativa de propriedades rurais, estradas e atrativos, o potencial classificado para a região fica entre médio e baixo com apenas algumas caracterizações de alto potencial.

## 5. CONCLUSÕES

O resultado obtido ao final do trabalho foi satisfatório e apresentou uma caracterização de potencialidades de ocorrência de incêndios interessante tendo em vista que não seguiu um padrão característico de incêndios florestais.

O mapa final proporciona um suporte importante na prevenção e combate aos incêndios florestais no parque, pois com o conhecimento prévio de quais são as regiões onde existe uma alta probabilidade que ocorra um incêndio, pode ser posicionado equipes de brigadistas ou monitoramento nestes pontos mais críticos, tornando o atendimento mais eficiente.

O processo de KDD é bastante amplo e abre diversas possibilidades de análise, pois permite a utilização de diferentes classificadores, algoritmos e variáveis. Para este trabalho, levando em consideração o resultado das estatísticas, pode-se dizer que a classificação por árvore de decisão e o algoritmo escolhido apresentaram bons resultados como esperado inicialmente.

É importante salientar que alguns refinamentos podem ser feitos para melhorar o modelo como, por exemplo, a utilização de uma escala maior para que variáveis como declividade e orientação da encosta tivessem um nível maior de detalhe. Outra opção seria a incorporação de mais variáveis ao modelo, uma vez que com uma quantidade maior de informações distintas o *software* conseguiria detalhar e caracterizar ainda melhor a área. A utilização de um método diferente de classificador juntamente com diferentes algoritmos também pode ser utilizado.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

- ADDRIANS, P. & ZANTINGE, D. **Data Mining**. Inglaterra: Addison-Wesley, 1996.
- ARONOFF, S.; **Geographical Information Systems: A Management Perspective**. Ottawa, WDI Publications, 1989.
- BATISTA, A. C. **Incêndios florestais**. Recife: Universidade Federal Rural de Pernambuco, 1990.
- BONTEMPO, G. C. **Ocorrência de incêndios florestais e educação ambiental nos parques abertos à visitação pública em Minas Gerais**. 2006. 162 p. Dissertação (Mestrado em Ciências Florestais) - Universidade Federal de Viçosa, Viçosa, 2006
- BROWN, A.A. & DAVIS, K.P. **Forest Fire – Control and Use**. New York, McGraw Hill, 2nd Ed., 1973. 686 p.
- CAMILO, C. O; SILVA, J. C. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**. Technical Report - RT-INF\_001-09 - Relatório Técnico August - 2009 – Agosto. Disponível em <<http://www.inf.ufg.br/biblioteca-digital>>. Acesso em: Out. 2013.
- CASTRO, D. M; MOURA, A. C. M. **PROCEDIMENTOS DE DATA MINING NA DEFINIÇÃO DE VALORES PARA AS ANÁLISES DE MULTICRITÉRIOS COMO APOIO À TOMADA DE DECISÕES E ANÁLISE ESPACIAIS URBANAS**. XXIV Congresso Brasileiro de Cartografia - Aracaju - SE - Brasil, 2010. Disponível em: <[http://www.arq.ufmg.br/SiteLabGeo/Laboratorio\\_Geo/2010.htm](http://www.arq.ufmg.br/SiteLabGeo/Laboratorio_Geo/2010.htm)>. Acesso em: Set. 2014.
- CERQUEIRA, M. **Utilização do geoprocessamento para estudo do uso e ocupação conflitante com a legislação ambiental na bacia hidrográfica do rio Manhuaçu, no município de Manhuaçu-MG**. 2006. 52 p. Dissertação (Graduação em Geografia) – Universidade Federal de Viçosa, Minas Gerais, 2006.
- CIOS, K. J; **Data Mining - A Knowledge Discovery Approach**. Springer, 2007.
- COCHRANE, M.A. **Fire science for rainforests**. Nature 42: 913-919. 2003.
- DIAS, M. M. **Parâmetros na escolha de técnicas e ferramentas de mineração de dados**. Portal Acta Scientiarum, Maringá, v. 24, n. 6, p. 1715-1725, 2002
- FAYYAD, U.M. et al. **Advances in Knowledge Discovery and Data Mining**. California: AAAI Press, 1996b.
- FAYYAD, U.M. et al. **The KDD Process for Extracting Useful Knowledge from Volumes of Data**. In: \_\_\_\_\_. **Advances in Knowledge Discovery in Data Mining**. Menlo Park: AAAI Press, 1996a.
- HAN, J; **Data Mining: Concepts and Techniques**. Elsevier, 2006

IBAMA - Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis. *Investigação de incêndios florestais / Alexandre de Matos Martins Pereira... [et al.]* – Brasília: Prevfogo/Ibama, 2011. 76 p.: il.; 16 cm.

IBAMA. **Plano de Manejo - do Parque Nacional da Serra da Canastra.** Brasília, 2005

IEF, Instituto Estadual de Florestas. Retirado de:

<http://www.ief.mg.gov.br/component/content/198?task=view>. Acessado em: 20/09/2014.

IEF, Instituto Estadual de Florestas; **Plano de Prevenção e Combate a Incêndios Florestais (PCIF)**, apostila, 2007.

MANNILA, H. **Data mining: machine learning, statistics and databases.** International Conference on Statistics and Scientific Database Management, Estocolmo, 8, 1996

MARTINS, F. T; DUTRA, V. L; FREITAS, C. C; FONSECA, R. F; GUIMARÃES, R. J. P. S; MOURA, A. C. M; SCHOLTE, R. G. C; AMARAL, R. S; DRUMMONT, S. C; FREITAS, C. R; CARVALHO, S. O. **Uso de árvore de decisão para predição de prevalência de esquistossomose no Estado de Minas Gerais, Brasil.** Anais XIII Simpósio Brasileiro de Sensoriamento Remoto, Florianópolis, 2007, INPE, p. 2841-2848.

MATOS, Alberto V. de M. **Estratégias de preservação de incêndios florestais o estado de Minas Gerais.** Congresso Brasileiro de Unidades de Conservação, 2002, Fortaleza. Anais, V.I, Conferências e Palestras, Fortaleza.

NOGUEIRA, G. S. et al.. **Escolha de locais para instalação de torres de detecção de incêndio com auxílio do SIG,** Revista da Sociedade de Investigações Florestais, v.26 n. 3, p.363-369, Viçosa, MG, 2002.

PEREIRA, A.A.; PEREIRA, L.C.; FERREIRA, R.V.; **Monitoramento dos Incêndios Florestais do Estado de Minas Gerais.** Disponível nos Anais XIII Simpósio Brasileiro de Sensoriamento Remoto, Florianópolis, Brasil.

PRUDENTE, T. D. **Geotecnologias aplicadas ao mapeamento de risco de incêndio florestal no Parque Nacional da Chapada dos Veadeiros e área de entorno.** 2010. Dissertação de mestrado em Geografia na Universidade Federal de Uberlândia, Uberlândia. 2010

RAMOS, P.C.M.; **Sistema Nacional de Prevenção e Combate aos Incêndios Florestais.** disponível nos Anais IPEF, 1995.

REZENDE, S. O.; **Sistemas inteligentes: fundamentos e aplicações.** Barueri, Manole, 2005.

SFERRA, Heloisa Helena; CORREA, Ângela M. C. Jorge. **Conceitos e Aplicações de Data Mining.** Jul/Dez de 2003, Revista Ciência & Tecnologia, PP. 19-34

SILVA, G. F. N. **Processamento de Imagens Digitais e Análise Espacial para o estudo da susceptibilidade a incêndios florestais nas regiões do Apaiaú e Ribeiro Campos-RR.** 2006 Dissertação de mestrado em recursos naturais Universidade Federal de Roraima, Boa Vista, 2006.

SOARES, R. V.; SANTOS, J. F.; **Perfil dos Incêndios Florestais no Brasil**. Disponível na Revista Floresta 2002.

WANG, J. **Encyclopedia of Data Warehousing and Mining**. Idea Group Reference, 2005









```

| | declividade <= 2.27: Media (5.0)
| | declividade > 2.27: Baixa (38.0/10.0)
dist_trilha > 6554.64
| dist_atrativ <= 20263.2
| | dist_trilha <= 7510.13
| | | declividade <= 19.18: Alta (20.0/3.0)
| | | declividade > 19.18
| | | | dist_atrativ <= 13242.8: Alta (4.0/1.0)
| | | | dist_atrativ > 13242.8: Media (8.0/2.0)
| | dist_trilha > 7510.13: Alta (35.0/7.0)
| dist_atrativ > 20263.2: Baixa (3.0)

```

Number of Leaves : 108

Size of the tree : 192

Time taken to build model: 0.12 seconds

==== Evaluation on training set ====

==== Summary ====

Correctly Classified Instances	693	73.0242 %
Incorrectly Classified Instances	256	26.9758 %
Kappa statistic	0.5719	
Mean absolute error	0.2437	
Root mean squared error	0.3491	
Relative absolute error	56.6065 %	
Root relative squared error	75.2411 %	
Total Number of Instances	949	

==== Detailed Accuracy By Class ====

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.726	0.058	0.796	0.726	0.759	0.926	Alta
	0.864	0.313	0.686	0.864	0.765	0.861	Media
	0.549	0.074	0.777	0.549	0.644	0.868	Baixa
Weighted Avg.	0.73	0.176	0.741	0.73	0.725	0.879	

==== Confusion Matrix ====

```

a b c <-- classified as
164 47 15 | a = Alta
24 362 33 | b = Media
18 119 167 | c = Baixa

```

### ANEXO 2 – MAPA DE POTENCIAL

